ArrayTrack$_{3.1.5}$ Demonstration

National Center for Toxicological Research
U.S. Food and Drug Administration
3900 NCTR Road,
Jefferson, AR 72079

ArrayTrack Overview
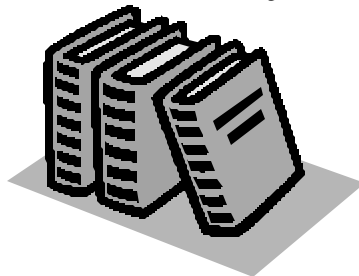
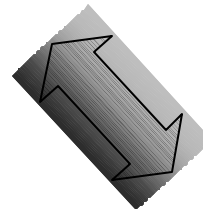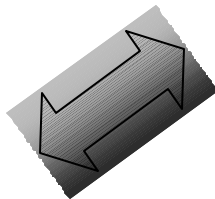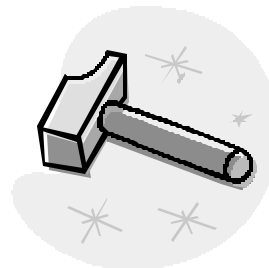## Microarray DB

Library

Tool

An integrated environment for microarray data management, analysis and interpretation.

Using ArrayTrack, the user can apply analysis method from Tool to microarray DB and then get Information from the linked LIB for biological interpretation
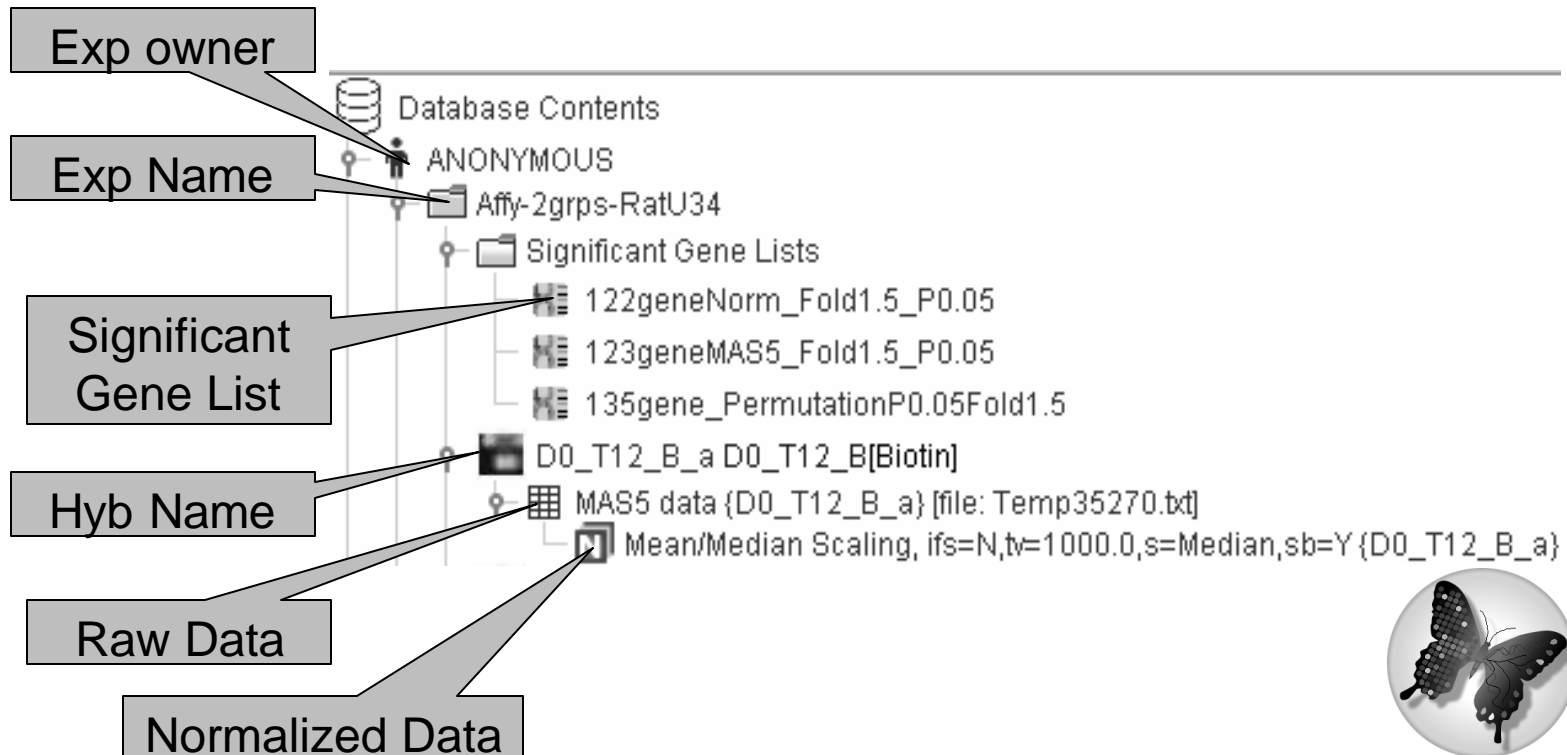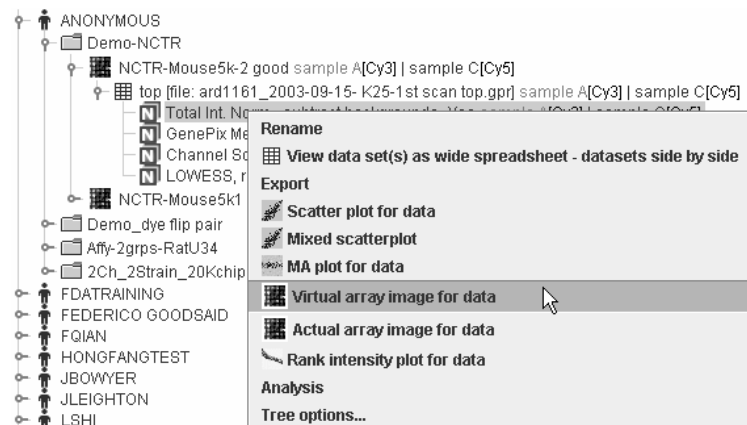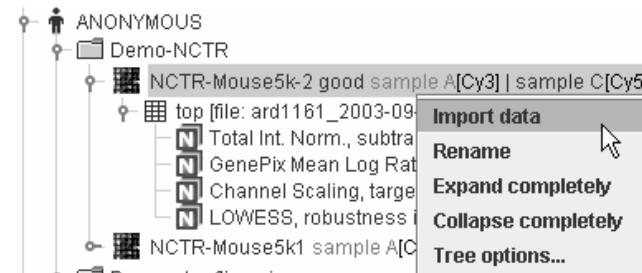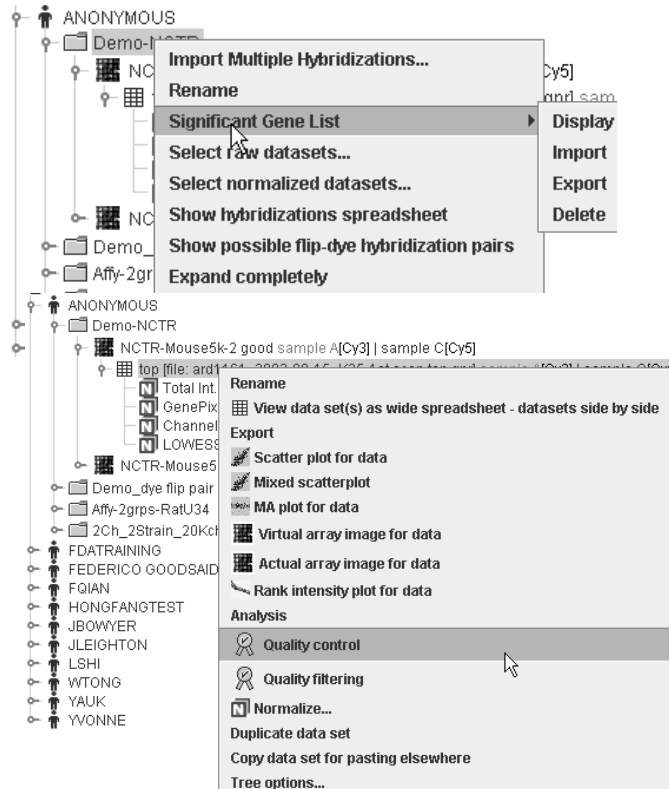
Microarray DB

Data is organized as a hierarchical tree structure:

Exploring DB

- Right-click exp, hybridization, raw data and normalized data to access various functions

Importing Data

1. create experiment
2. create hybridization
3. create array type (if not existed in AT Chip Lib)
4. create sample

MicroarrayDB stores DNA microarray gene expression experimental data together with essential annotation information about an experiment, its protocol and the samples.

Support data from one-channel and two-channel.

Searching Libraries

Library
- ID Converter
- Gene Library
- Pathway Library
- Protein Library
- IPI Library
- Orthologene Library
- GOFFA Library
- Chip Library
- Toxicant Library
- EDKB Library

There are nine libraries in ArrayTrack.
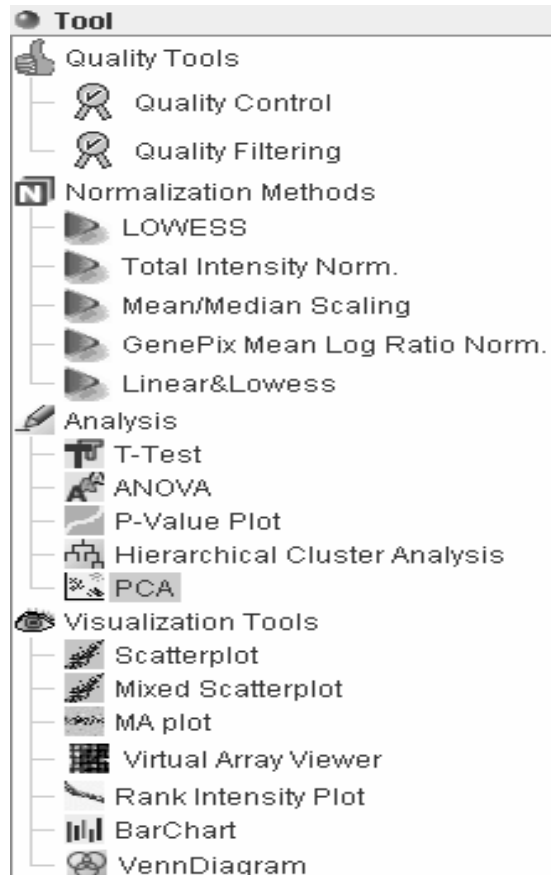All the libraries are interlinked.

The libraries integrate gene, protein, pathway and other data allowing data interrogation and mining of data across data types.

Several ways to activate these libraries
-From library panel
-From the Library pull-down menu
-From the results of analysis (e.g. T-test)

Analysis Tools

Tools:

•Quality Tools

Provides various visual plots and numerical parameters for measuring the quality of a hyb, and filtering the unwanted spots.

•Normalization Methods

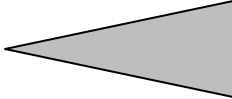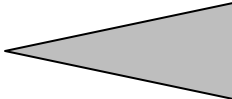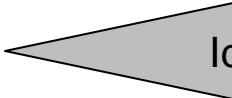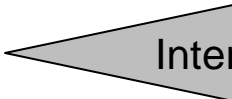Correct systematic variations in microarray data introduced by experimental factors

•Analysis

Provides different methods to evaluate the microarray data

•Visualization Tools

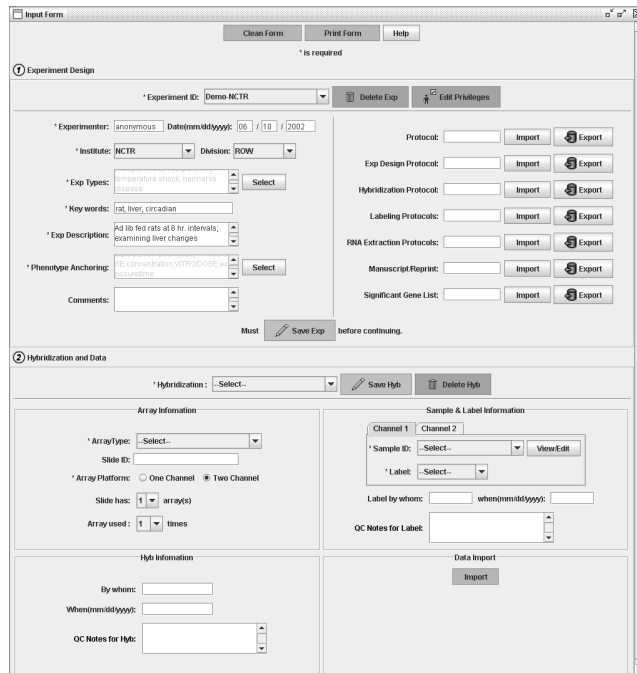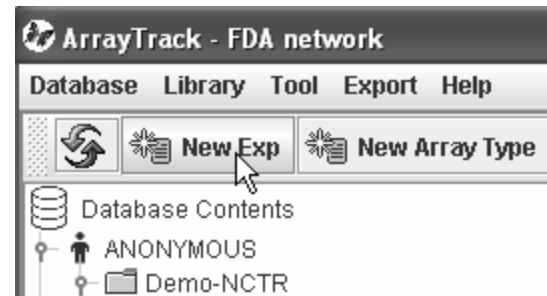Provides a direct view to identify abnormalities within data.

Microarray Data Processing

- Importing data — Loading array data into AT
- Normalization — Ensure cross-chip comparison
- Gene Selection — Identify a list of significant genes
- Interpretation — Interpret data using pathways and GO
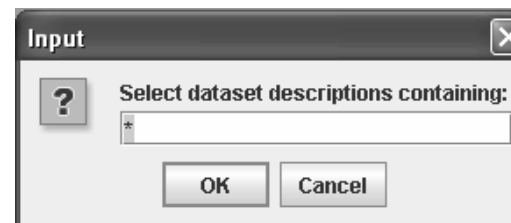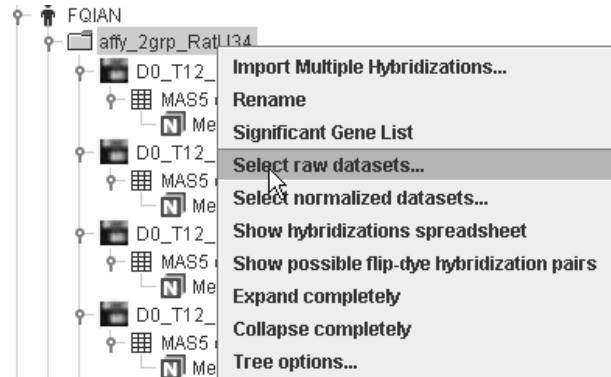
Array Track

Importing Data

1. create experiment
2. create hybridization
3. create array type (if not existed in AT Chip LIB)
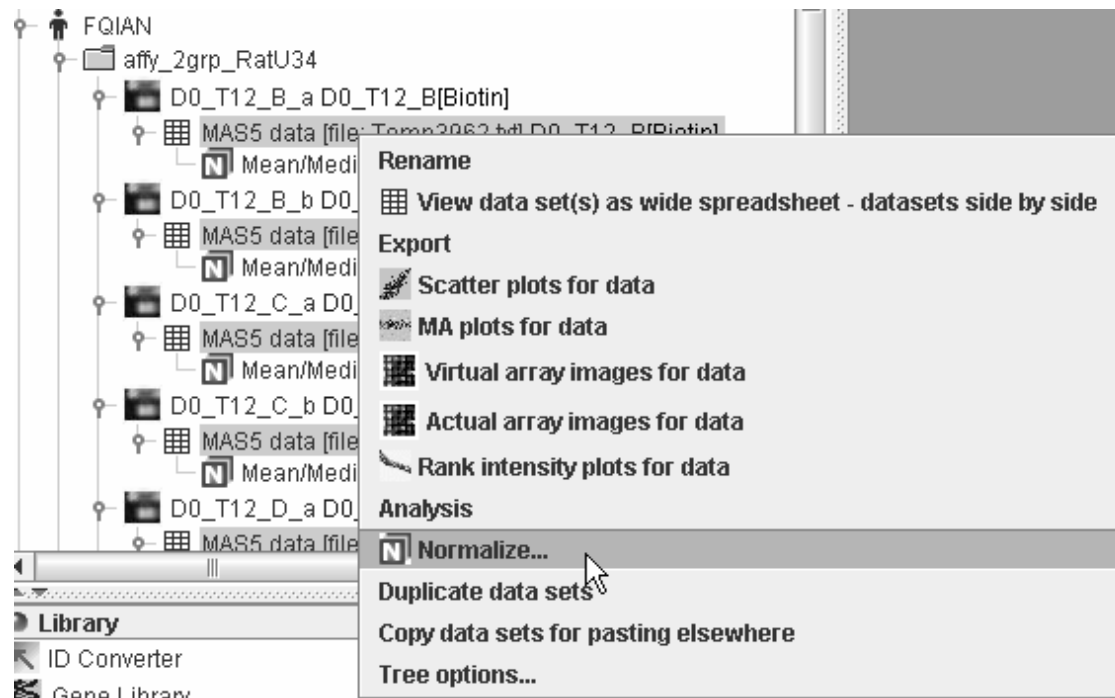4. create sample

Normalization

- To remove systematic variations across chips and ensure a valid cross-chip comparison
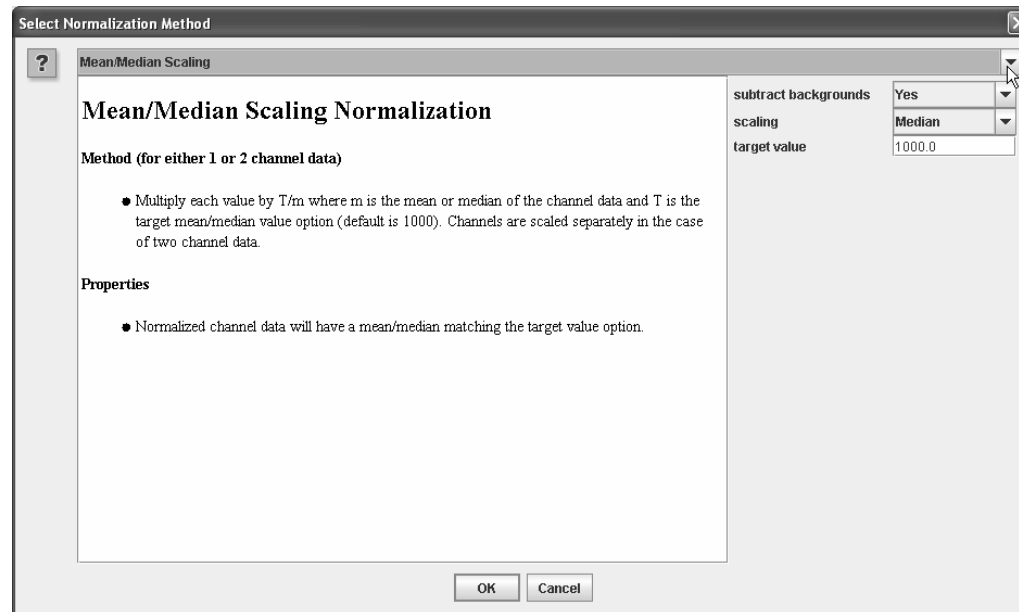
1. Right-click an experiment
2. Select "raw datasets…"

Normalization (-continued)

3. Right-click any highlighted raw data

4. Select "Normalize…"

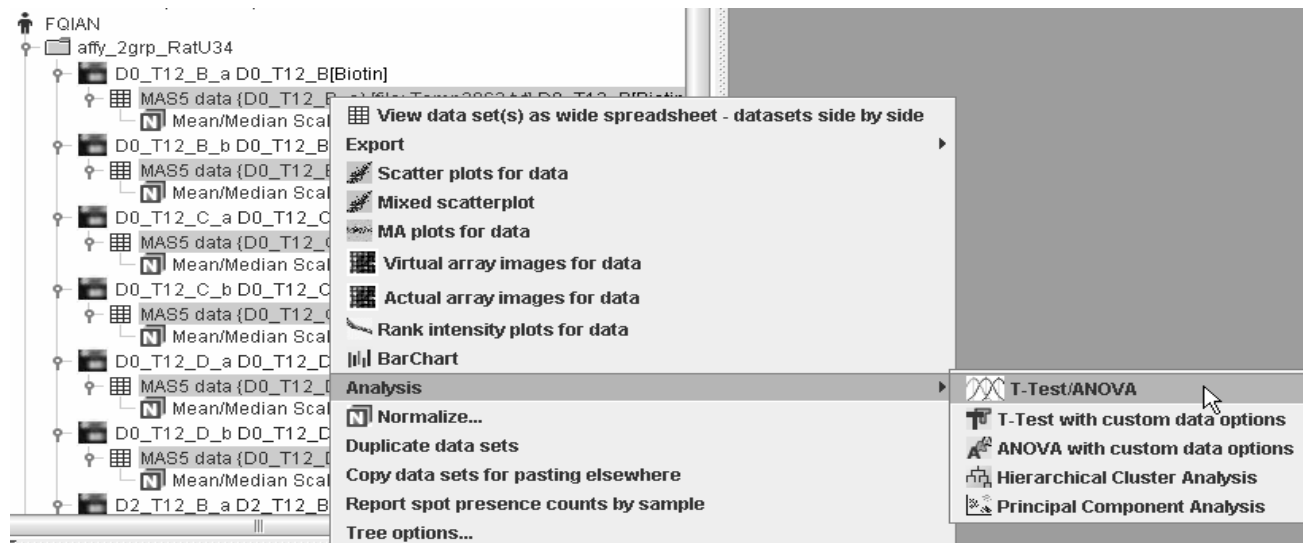# Normalization (-continued)

- Choose normalization method

- For Affy data, choose "Mean/Median Scaling"
- For 2-channel data, the default method is "Lowess"

Gene Selection

- Highlight and Right-click the experiment.
- Select "Normalized datasets…" and click "OK"
- Right-click the highlighted normalized data
- Choose "Analysis->T-test"

Gene Selection (-continued)

Assign the data into 2 groups
-different dose
-different time
-or different animal

Gene Selection (-continued)

Parameters used to filter the genes to get significant gene list

Access other analysis methods like HCA, PCA, Volcano plot, etc.

Gene Selection (-continued)
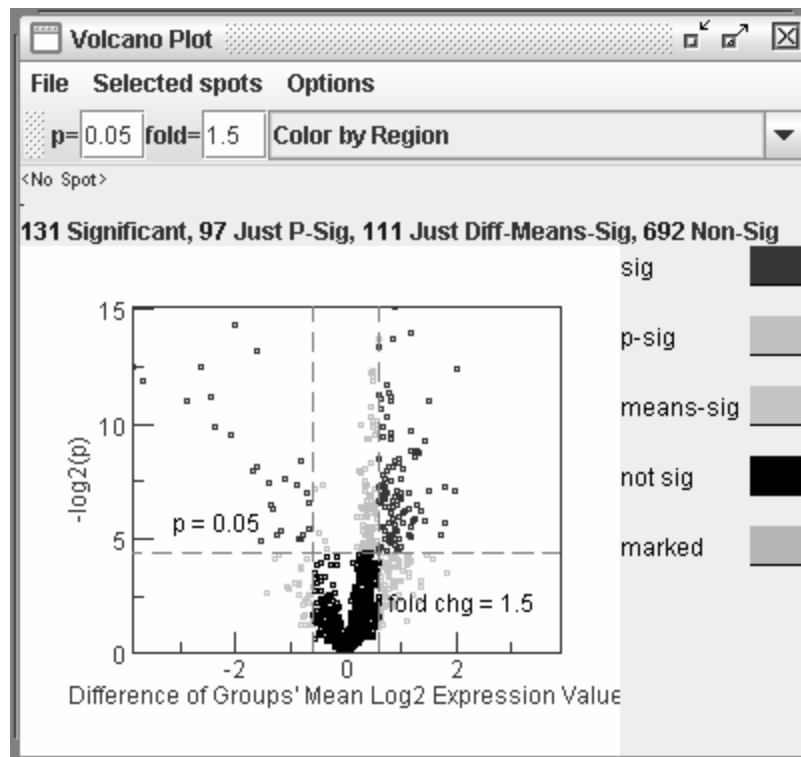
Volcano Plot

Significant Gene List

- Create
- Display
- Import
- Export
- Delete

Significant Gene List(-cont)

- Create significant gene list

  From T-test/Anova result

  set filter criteria

  e.g. P-value, fold change

  to get significant gene list

Create significant gene list

Filter criteria

Significant Gene List(-cont)

- **Display/Import/Export/Delete Significant Gene List**

  Right-click the experiment name

  select Significant Gene Lists

  select Display

  Import

  Export

  Delete
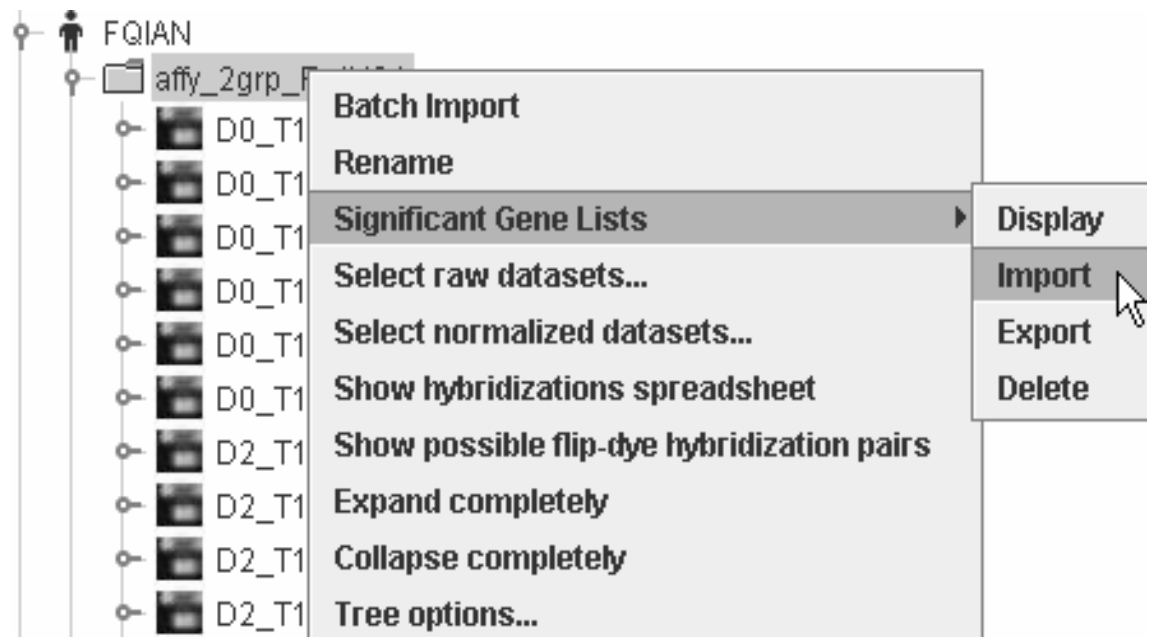
Significant Gene List(-cont)

- **VennDiagram -** get common genes from 2~3 significant gene lists

Interpretation

Link the significant genes to Gene Library for data interpretation

Interpretation (continued)

- The significant genes are listed here in Gene Library.

- Can search and sort the Gene library

- There are links to other Libraries(Kegg, Pathart)

Interpretation (continued)

- KEGG – Kyoto Encyclopedia of Genes and Genomes
  http://www.genome.jp/kegg/
- KEGG is a suite of databases and associated software.
- KEGG Pathway database provides the information of metabolic, regulatory and disease pathways; Most of them are metabolic pathways.

Interpretation (continued)

PathArt (Jubilant) – a pathway database
- The Pathways (over 600 mammalian disease and signaling)
- The Pathways is a collection of manually curated information from literature and public domain databases.

In ArrayTrack

	Human	Rat	Mouse
Kegg	134	116	124
PathArt	587	151	297

Interpretation (continued) Kegg

Genes involved in a pathway

Pathway name

Pathway category

Statistical significance of the pathway

Kegg Pathway

Interpretation (continued) Kegg

Double-click a specific pathway, the pathway map will be displayed and the genes involved in the pathway are highlighted.
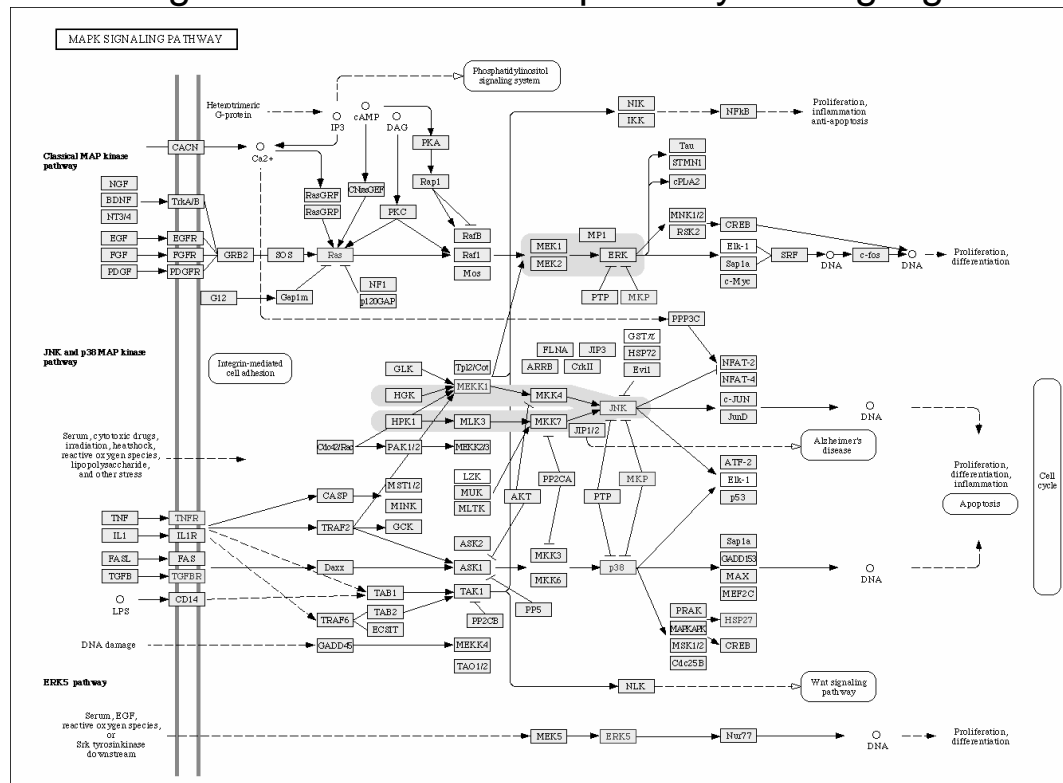
Interpretation (continued) PathArt

Genes   Pathways   Physiology/disease   Statistical significance of the pathway

| Component n...	Pathway Name	Organism	Physiology/...	IS_PHYSIO...	Pathway Ty...	Fisher P V...
Mapk14	CCR5 Mediated Pathway	Mus muscu...	Acquired I...	Disease	Signaling	0.089682
Gclc	Tat Signaling Pathway	Mus muscu...	Acquired I...	Disease	Signaling	0.236507
Mapk14	AGEs Signaling Pathway	Mus muscu...	Alzheimers	Disease	Signaling	0.149045
Mapk14 Tnfrsf1a	Amyloidbeta-peptide Signal...	Mus muscu...	Alzheimers	Disease	Signaling	0.118019
Tnfrsf1a	IFN Signaling Pathway	Mus muscu...	Alzheimers	Disease	Signaling	0.634464
Mapk14	LPS Signaling Pathway	Mus muscu...	Alzheimers	Disease	Signaling	0.386674
Map3k1	DR4/5 Signaling Pathway	Mus muscu...	Apoptosis	Physiology	Signaling	0.064863
Mapk14	IGF Signaling Pathway	Mus muscu...	Apoptosis	Physiology	Signaling	0.717471
Ngfr	NGF Signaling Pathway	Mus muscu...	Apoptosis	Physiology	Signaling	0.204750
Mef2d	T-cell apoptosis	Mus muscu...	Apoptosis	Physiology	Signaling	0.149045
Mapk14 Tgfbr2	TGF Beta Induced apoptosis	Mus muscu...	Apoptosis	Physiology	Signaling	0.051402
Tnfrsf1a	TNF Signaling Pathway	Mus muscu...	Apoptosis	Physiology	Signaling	0.776034
Mapk14	TNFR1 Signaling Pathway	Mus muscu...	Apoptosis	Physiology	Signaling	0.246824
Adprt	TRAIL Mediated Apoptosis	Mus muscu...	Apoptosis	Physiology	Signaling	0.378170
Mapk14	UV induced Antiapoptotic P...	Mus muscu...	Apoptosis	Physiology	Signaling	0.113880
Cd44	WNT Signaling Pathway	Mus muscu...	Apoptosis	Physiology	Signaling	0.644709
Mapk14	Interleukin Signaling Pathw...	Mus muscu...	Atheroscle...	Disease	Signaling	0.926753
Met	HGF Signaling Pathway	Mus muscu...	Breast Ca...	Disease	Signaling	0.064863
Tgfbr2	TGF Signaling Pathway	Mus muscu...	Breast Ca...	Disease	Signaling	0.267067
Cd44	CD44 Signaling Pathway	Mus muscu...	Cell Adhe...	Physiology	Signaling	0.052216
Tp53	ATR-ATM Signaling Pathway	Mus muscu...	Cell Cycle	Physiology	Signaling	0.204750
Map3k1 Vegf	Bcr-Abl Signaling Pathway	Mus muscu...	Chronic M...	Disease	Signaling	0.217718

Total genes =60, Found genes= 17, Found pathways= 58

Interpretation (continued)   PathArt
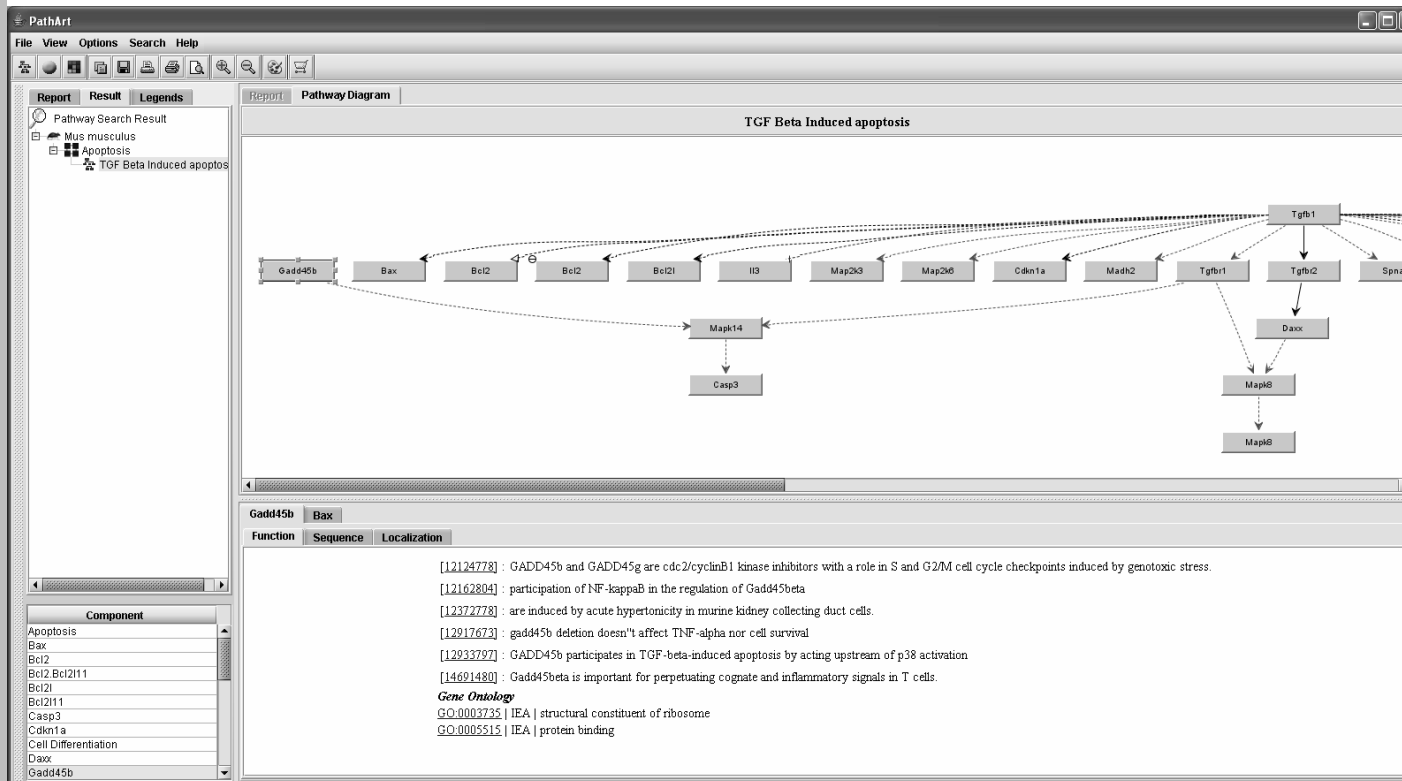
Interpretation (continued) GOFFA

- Developed based on Gene Ontology(GO) database
- Grouping the genes into functional classes
- GO- three ontologies

-Molecular function: activities performed by individual gene products at the molecular level, such as catalytic activity, transporter activity, binding.

-Biological process: broad biological goals accomplished by ordered assemblies of molecular functions, such as cell growth, signal transduction, metabolism.

-Cellular component: the place in the cell where a gene product is found, such as nucleus, ribosome, proteasome.

Interpretation (continued)   GOFFA

Data Exploring

- Scatter Plot

  plot the fluorescence intensity data of Cy3 vs Cy5 for the same array

- Mixed Scatter Plot

  compare two arrays in one plot, applies to both 2-channel and 1-channel data.

- Principal Component Analysis (PCA)

  PCA is a way of identifying the data patterns and highlighting the data's similarity and difference.

- Bar Chart

  displays expression data for a single gene across multiple arrays within the same experiment or across different experiment.

- HCA

  is comprised of agglomerative methods and divisive methods that finds clusters of observations within a data set.

Data Exploring   scatter plot
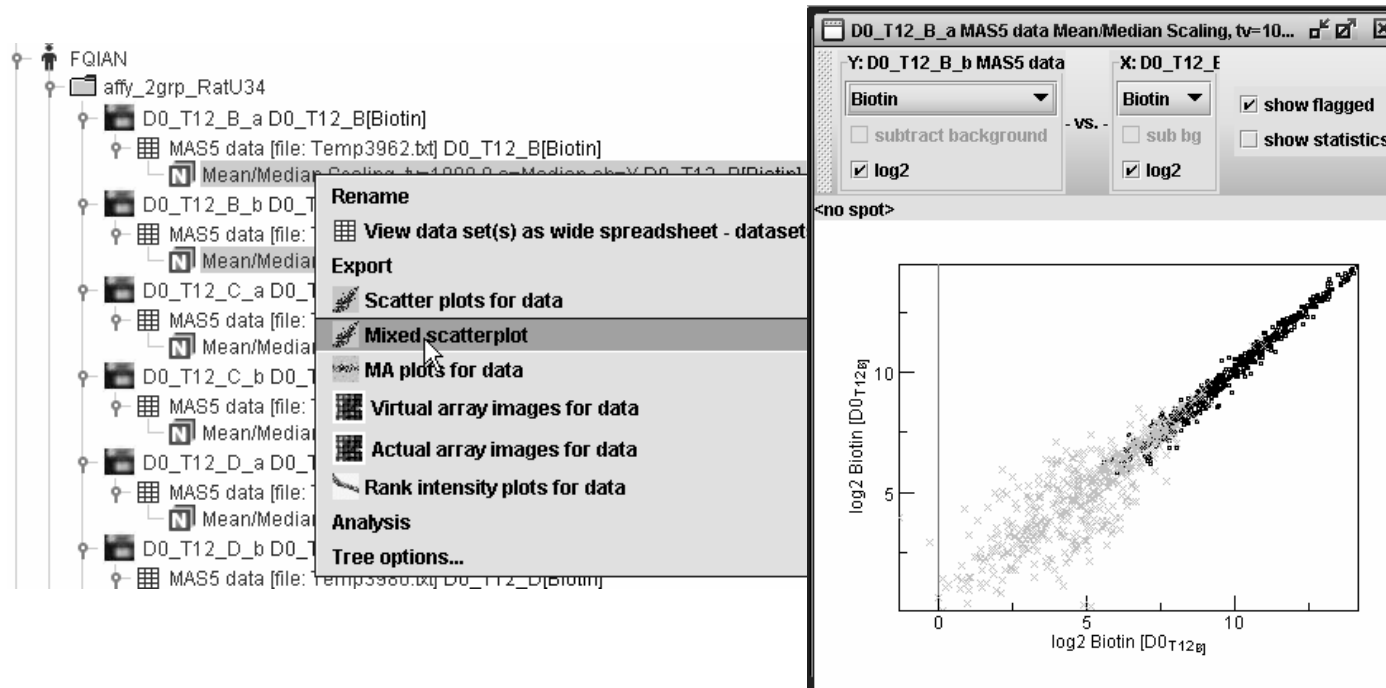
Data Exploring   Bar chart

Access Bar chart from T-Test results.

Bar chart – display expression data for a single gene across multiple arrays in the same experiment or across different experiments.

Data Exploring   Bar chart   continued
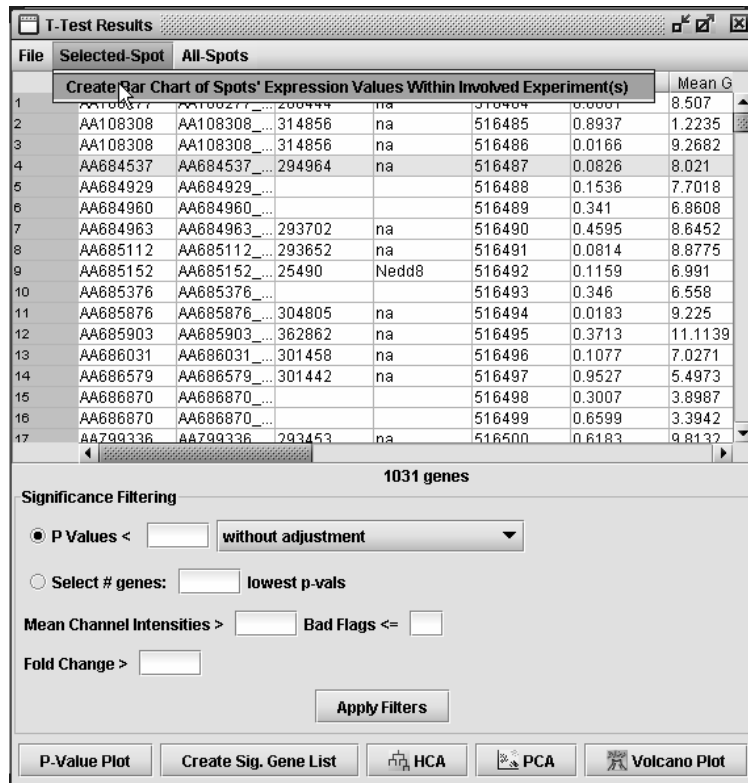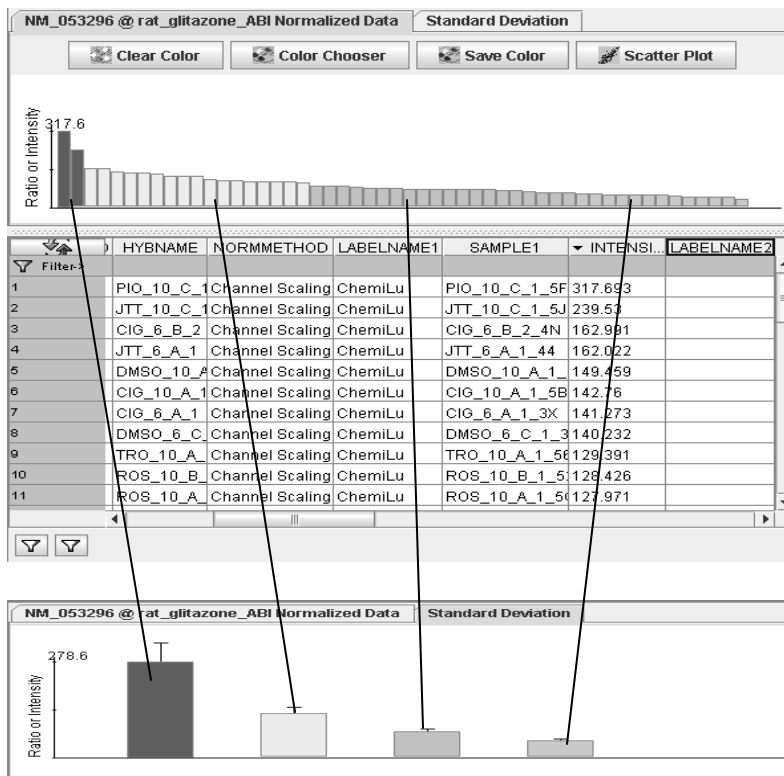

Grouping multiple arrays marked in different colors

Standard deviation bar chart for the above groups. The bar height represents the mean intensity, while the T-line above the bar stands for the value of SD. The color for each bar echoes the color of the bar chart at the top, in the same order.

Data Exploring PCA

- Commonly used before gene selection
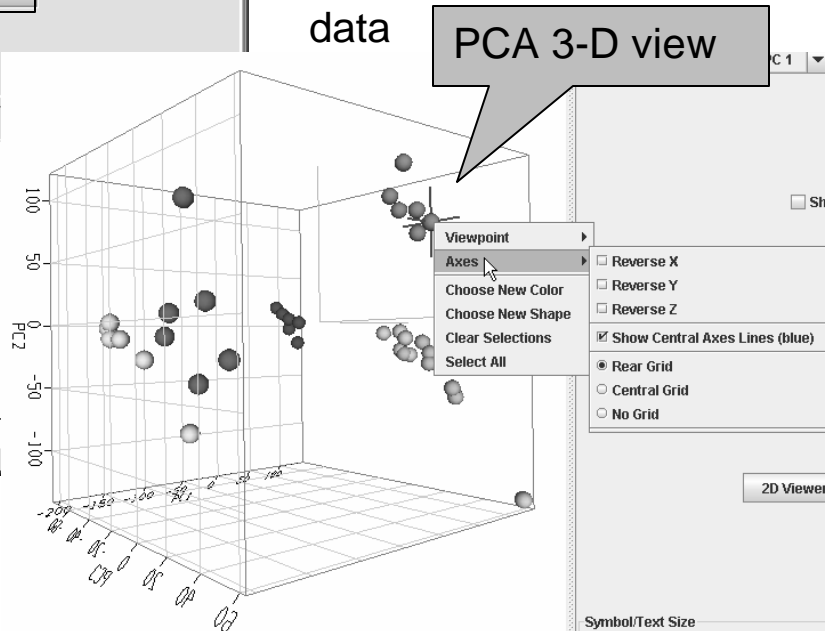- To investigate the inter-sample relationship based on the gene expression profile
- Identify the biological/technical replicates
- View the variance of a multidimensional data

PCA 3-D view

Data Exploring   HCA
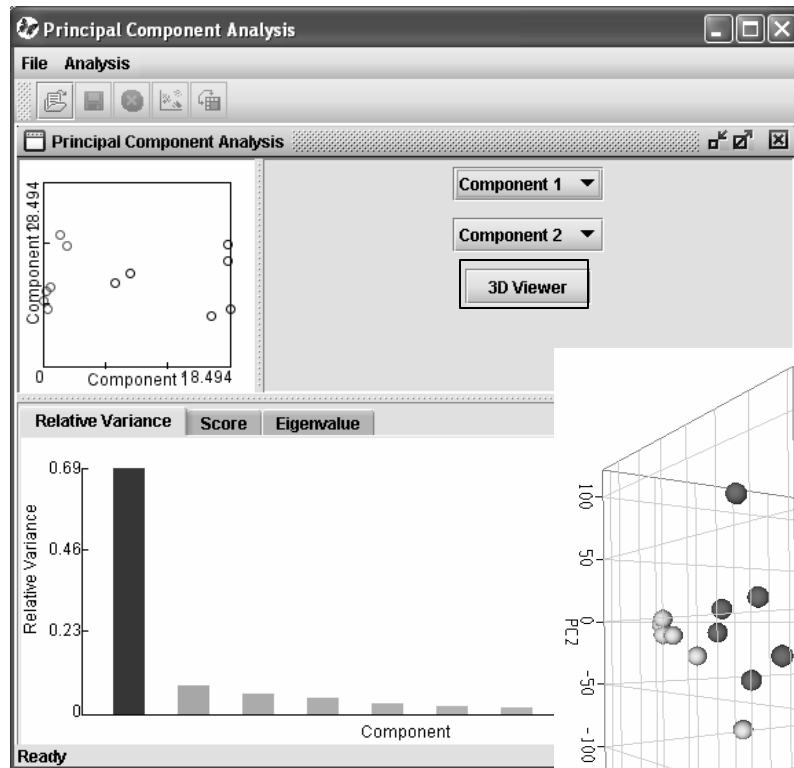
Accessing ArrayTrack

- FDA Internal:
  http://weblaunch.nctr.fda.gov/jnlp/arraytrack/index.html

  http://weblaunch.nctr.fda.gov/jnlp/arraytrack/citrix/index.html

- FDA External:
  http://edkb.fda.gov/webstart/arraytrack/index.html

Technical Support

NCTRBioinformaticsSupport@nctr.fda.gov

ArrayTrack is developed by the U.S. Food and Drug Administration, National Center for Toxicological Research (FDA/NCTR).

FDA/NCTR reserves all rights for the software .

Thank you!

National Center for Toxicological Research
U.S. Food and Drug Administration